REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	4 August 1998	

4. TITLE AND SUBTITLE
DETERMINING SPECIES COMPOSITION USING TEMPORAL NDVI TRAJECTORIES DERIVED FROM SATELLITE REMOTE SENSING MEASUREMENTS

5. FUNDING NUMBERS

6. AUTHOR(S)
Matthew J. Tracy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Kansas State University

8. PERFORMING ORGANIZATION REPORT NUMBER
98-031

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

14. SUBJECT TERMS	15. NUMBER OF PAGES
	31
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

Determining Species Composition Using Temporal NDVI Trajectories
Derived From Satellite Remote Sensing Measurements

by

Matthew J. Tracy

B.S., U.S. Air Force Academy, 1996

A Report

submitted in partial fulfillment of the

requirements for the degree

MASTER OF ARTS

Department of Geography

College of Arts and Sciences

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1997

Approved by:

Major Professor

Abstract

Significant distinctions in phenological properties exist between plant communities. These differing characteristics can be the result of varying plant types, environmental stressors, or both of these factors. This paper explores the ability of satellite remote sensing to determine relative amounts of C_3 and C_4 vegetative lifeforms contained within a heterogeneous canopy based on their unique phenological qualities. Changes in the structure and dynamics of an environment, in this case a tallgrass prairie, can be indicators of even larger stresses on the ecosystem. The C_3 and C_4 plants of a tallgrass prairie, for example, have been shown to be particularly sensitive to environmental changes. The ability to distinguish between these lifeforms can be complex due to the fact that they appear spectrally similar at a single point in time. Other studies have proven this to be possible using hand held, close range remote sensing measurements with fine spatial resolution. However, the ability to distinguish these lifeforms using imagery of relatively coarse resolution has not yet been explored. Twenty-six SPOT images from March to November of 1987 were obtained for analysis. Percentages of C_3 and C_4 plants were determined using discriminant function mixture models based on metrics derived from the temporal trajectory of the Normalized Difference Vegetation Index (NDVI). The percentages derived from satellite measurements were compared with data collected on the ground. Classification accuracies between 50%-60% were obtained using the techniques discussed in the paper. The

results presented in this report seem to be a promising indicator that determining amounts of C_3 and C_4 within a canopy is possible using satellite remote sensing.

Abstract

Significant distinctions in phenological properties exist between plant communities. These differing characteristics can be the result of varying plant types, environmental stressors, or both of these factors. This paper explores the ability of satellite remote sensing to determine relative amounts of C_3 and C_4 vegetative lifeforms contained within a heterogeneous canopy based on their unique phenological qualities. Changes in the structure and dynamics of an environment, in this case a tallgrass prairie, can be indicators of even larger stresses on the ecosystem. The C_3 and C_4 plants of a tallgrass prairie, for example, have been shown to be particularly sensitive to environmental changes. The ability to distinguish between these lifeforms can be complex due to the fact that they appear spectrally similar at a single point in time. Other studies have proven this to be possible using hand held, close range remote sensing measurements with fine spatial resolution. However, the ability to distinguish these lifeforms using imagery of relatively coarse resolution has not yet been explored. Twenty-six SPOT images from March to November of 1987 were obtained for analysis. Percentages of C_3 and C_4 plants were determined using discriminant function mixture models based on metrics derived from the temporal trajectory of the Normalized Difference Vegetation Index (NDVI). The percentages derived from satellite measurements were compared with data collected on the ground. Classification accuracies between 50%-60% were obtained using the techniques discussed in the paper. The

results presented in this report seem to be a promising indicator that determining amounts of C_3 and C_4 within a canopy is possible using satellite remote sensing.

Table of Contents

List of Tables

Table

Acknowledgments

There are a number of people who have made a significant impact on my academic development at Kansas State University and who deserve recognition. I would first like to thank Dr. Doug Goodin who has spent considerable time advising me on many issues, including the development of this report. I would like to thank my other committee members, Dr. John Harrington and Dr. Lisa Harrington, for their insights and suggestions. Thanks also to all the other faculty, staff, and students who have combined to make my experience with the Department of Geography a memorable and rewarding experience.

I would also like to recognize the Division of Biology for providing me with certain data sets used in this project. Specifically, data set PVC02 was supported by the NSF Long Term Ecological Research Program at Konza Prairie Research Natural Area.

Introduction

The use of remote sensing by organizations, academia, and individuals has exploded over the course of the last several decades. In the earliest stages of development, few people realized the benefits that remote sensing had to offer. This technology has advanced rapidly during the last 20 years and is now widely used in a number of professions. Satellite sensors are useful because they can provide detailed information about rather large geographic areas. Satellite images are routinely used to obtain thematic information about the earth's surface such as temperature, vegetative cover, and soil moisture.

One of the more common applications of remotely sensed images is to obtain information about various vegetative parameters such as biomass or vigor. This information can be used as an indicator of the response of the biosphere to a variety of earth system disturbances (Fischer, 1994). Determination of vegetative parameters, though, can be a complex process. It can be difficult to derive some of these parameters from coarse resolution satellite data, or when trying to extract them from a heterogeneous vegetated canopy. The aim is often to extract information on a smaller scale than the scale of observation. This is an area of ongoing research in the field of remote sensing, and

will be a primary focus of this paper as well.

The objective of this study is to determine relative amounts of C_3 and C_4 vegetative lifeforms contained within a heterogeneous canopy in a tallgrass prairie environment using satellite imagery. Percentages derived from satellite measurements will be compared with data collected on the ground. The C_3 and C_4 lifeforms can be discriminated based on their inherent phenological differences throughout the growing season. In Northeast Kansas C_3 species tend to be invaders to a tallgrass prairie and thrive in the cooler spring and fall months, whereas C_4 species are generally native to a tallgrass prairie and thrive during the warmer summer months (Goodin and Henebry, 1997).

The unique photosynthetic pathways associated with C_3 and C_4 vegetative lifeforms are manifested by physical differences throughout the growing season. The photosynthetic pathway describes a series of processes that convert carbon dioxide into energy in plants. Waller and Lewis (1979) have identified C_3 and C_4 lifeforms as being the two most important photosynthetic pathways.

Five characteristics have been widely used to classify vegetation according to the photosynthetic pathway. These include: (1) photosynthetic products, (2) CO_2 compensation and photorespiration, (3) oxygen suppression, (4) leaf

anatomy, and (5) carbon isotope discrimination (Waller and Lewis, 1979). The specifics of these characteristics will not be expanded on for the purposes of this project. Waller and Lewis (1979) present a detailed discussion of this matter. It is simply important to note that C_3 and C_4 lifeforms have some unique properties.

In addition to these five characteristics, C_3 and C_4 vegetation can be discriminated based on characteristics of the temporal trajectories of the Normalized Difference Vegetation Index (NDVI) provided by satellite imagery. Vegetation that exhibit a C_3 photosynthetic pathway generally green up, or are more vigorous, during both the early and later parts of the growing season. Vegetation characterized by the C_4 pathway is more vigorous in the warmer mid-summer months. The relative amount of vegetative vigor can be estimated using the NDVI. On examination of several NDVI trajectories from the data in this project, it appears that sites with low C_3 content also reach higher NDVI values than those with high C_3 content. Transects with low C_3 content seem to reach certain percentages of cumulative NDVI earlier in the year as well.

The marked differences between C_3 and C_4 plants can help to explain aspects of structure in terrestrial ecosystems and the importance of warm season and cool season

plant classification in range management (Waller and Lewis, 1979). It is important to monitor the changes that occur in a tallgrass prairie environment. The same holds true for other ecosystems as well. By understanding the changes that take place in these particular environments, scientists and resource mangers will be able to care for them more effectively. Changes in the structure and dynamics of an environment, in this case a tallgrass prairie, can be indicators of even larger stresses on the ecosystem. The C_3 and C_4 vegetative lifeforms of a tallgrass prairie, for example, have been shown to be particularly sensitive to environmental changes.

Literature Review

A critical aspect of this report are the phenological characteristics of C_3 and C_4 lifeforms. "When planning to study phenology with satellite sensors, the questions of resolution and sampling design become apparent" (Fleischman and Walsh 1991). It is very important that the spatial, spectral, radiometric, and temporal resolution of the sensor are sufficient to detect the sometimes subtle changes of a vegetative canopy over the course of a growing season. This is critical if classification of plant species is based on differing phenology. Collins (1978) used narrow band

airborne instrumentation to sense phenological responses. Collins was effectively able to show phenological change between different species (Fleischmann and Walsh, 1991). It is this ability to distinguish among different vegetation types that has been relied upon in other studies.

The use of vegetation indices to detect phenological changes has received increased attention in recent years. The objective of vegetation indices is to derive information on the nature and state of vegetation by using various wavelengths selected to provide a strong signal from the vegetation, and at the same time contrast with background elements such as soil (Malingreau, 1989). The NDVI is one of the more widely used vegetation indices. It is defined as the ratio of the difference between the near infrared and red reflectances to their sum, (NIR-R)/(NIR+R). This particular vegetation index has the benefit of normalizing the differences in spectral reflectance.

There have been numerous studies that attempt to relate seasonal NDVI trajectories to cover type. Many of these studies deal with the problem of accurately classifying vegetation types that may appear to be spectrally similar at a single point in time, but are distinguishable based on their seasonal trajectories of NDVI, or other vegetation index (Wiegand and Richardson 1987; Fleischamn and Walsh

1991; Kremer and Running 1993; Fischer 1994; Lascassies et al. 1994; Reed et al. 1994; DeFries et al 1995; Goodin and Henebry, 1997). Using NDVI alone has not always proven to be the most effective technique for distinguishing among vegetation.

Other studies examine the relation between different vegetation indices and their ability to distinguish among cover types. It has been shown that, in some circumstances, a Relaxation Vegetation Index (RVI) is a better predictor of percentage plant cover than NDVI alone (Zhuang et. al., 1993). Although the ability of both the NDVI and RVI were found to be reasonably good predictors of percentage plant cover, RVI did prove to be a slightly better indicator in this particular study.

Oleson (1995) present a technique for extracting subpixel cover type reflectances from the mixed pixels of coarse spatial resolution data. They use weights representing the proportions of cover types within the mixed pixels and spectral band reflectances in multiple linear regression analysis to extract mean cover type reflectances. They determined that the accuracy of retrieved reflectances is most sensitive to errors in the coarse resolution data and least sensitive to errors in the weights.

DeFries, Hansen, and Townshend (1995) examined the use

of metrics derived from the NDVI temporal profile, as well as metrics derived from observations in red, infrared, and thermal bands to improve discrimination between 12 different cover types. They determined that some of the best metrics for discriminating cover types were mean NDVI, maximum NDVI and NDVI amplitude.

Reed and others (1994) also developed a series of metrics that were used to used to discriminate among cover types. They developed 12 metrics which are closely tied to key phenologic events. These measures include the onset of green up, time of peak NDVI, maximum NDVI, rate of green up, rate of senescence, and integrated NDVI. Their analysis showed a strong relation between the satellite derived metrics and predicted phenological characteristics.

Other investigators have also used an integrated NDVI approach, and have proved that this profile is a rather strong predictor of cover type. The integrated approach actually proved more accurate than the non-integrated technique (Kremer and Running, 1993; Henebry and Goodin, 1997). The advantage of using an integrated NDVI curve is that it enhances both minor seasonal differences, as well as overall variation among different vegetative signatures (Kremer and Running 1993). Generally the NDVI in any of the various forms has shown to be effective for distinguishing

among cover types in a heterogeneous canopy, or within a pixel. An integrated NDVI approach will be used for this report based on the indication that this technique can more effectively distinguish vegetation than NDVI alone.

Study Area

The Konza Prairie Research Natural Area (KPRNA) is a 3487 ha area located about 10km south of Manhattan, Kansas (Henebry 1993). This area is operated in part by Kansas State University and is used extensively for research by scientists and students. The Konza Prairie is in the Flint Hills, a narrow band of rolling hills that stretch from near the northern border of Kansas, south into Oklahoma. Henebry (1993) notes that erosion of underlying Permian limestone and shale sediments formed the hills, which are characterized by steep slopes, and flat ridges. The rugged topography of the area discouraged any widespread cultivation, and the vegetation is predominately native. The vegetation is dominated by big bluestem, little bluestem, indiangrass, and switchgrass. Oak and elm trees cover about 6% of KPRNA and are primarily found along stream channels (Henebry 1993).

The climate in the region is fairly typical of the Central Great Plains. Summers are rather hot, and winters

can be very cold. Winds throughout the year can be strong. The average maximum and minimum temperatures for January are around 8.0 C and 3.0 C, and for July are near 33.0 C and 20.0 C (Abrams and Hulbert 1986). Average precipitation is 835 mm, with approximately 70% occurring during the warmest 6 months (Abrams and Hulbert 1986). Annual precipitation, however, is highly variable from year to year.

A fire management plan was introduced across the KPRNA in 1971 (Gibson and Hulbert 1987). This plan placed watersheds under burning schedules of 1,2,4, and 10 year intervals. Other watersheds, however, are left unburned. In October of 1987, bison were reintroduced into 5 watersheds totaling approximately 450 ha (Henebry 1993).

Data Description and Quality:

A series of 26 satellite images from the growing season of 1987 were acquired for analysis. These images were used to produce temporal trajectories of NDVI for 72 transects located throughout the Konza Prairie. Relative amounts of C_3 and C_4 vegetative lifeforms will be determined based on metrics derived from these trajectories. SPOT scenes ranging from March to November were obtained from the First ISCLP Field Experiment (FIFE) compact disk #5. FIFE was an in-depth study conducted in a 15 square kilometer area south

of Manhattan from 1987 to 1989. The underlying purpose of the research, which was funded by NASA, was to determine the feasibility of using remotely sensed data to collect information about biophysical properties of the earth's surface. This study produced volumes of data which are available to the public.

All of the images obtained from the FIFE disk have been extensively preprocessed using a number of techniques. All scenes have been atmospherically corrected, converted to reflectance, transformed to NDVI, and registered to ground control points. Required images were simply decompressed from the disk having already been converted to NDVI and registered to the UTM coordinate system. All images were resampled from 20 meter spatial resolution to 30 meter resolution when they were georeferenced. A complete description of all preprocessing procedures can be accessed from the FIFE on-line documentation included with the compact disk.

The northwest corner of the images corresponds to UTM coordinates of 4,334,000 Northing and 705,000 Easting in UTM zone 14. The FIFE information system staff conducted quality checks throughout the production process. According to the FIFE staff, the raw data selected for processing is of generally high quality. Independent estimates by members

of the FIFE staff have established that the correcting algorithms can produce reflectance values accurate to 1% absolute. The final NDVI product presented on the disk is therefore a relatively accurate representation of the surface state, both radiometrically and spatially.

Other data utilized for this project include vector and raster GIS coverages of the Konza Prairie. The vector GIS coverage contained information on each individual watershed located within the Konza Prairie. The watersheds are distinguished based on the experimental treatments utilized by researchers. Treatments include variable burning, grazing, and nitrogen applications. The vector GIS coverage was used to identify watersheds of particular interest. Transects were located in watersheds 001C, 020B, N20B, N01B, 001D, N04D, 004B, 004D, 0004F, 002D, and 002C. A 30 meter resolution raster layer was utilized to further stratify the watersheds into uplands and bottomlands.

Species composition data from the 72 transects located throughout the area was an additional source of information. This is very important for two primary reasons. First, a subset of this information is used to generate classification coefficients that will be used to classify unknown cases. Secondly, these data were used to develop an accuracy assessment of the final results of the

classification. All species composition data were recorded by personnel of the Division of Biology at Kansas State University. The purpose of the study conducted by the Division of Biology was to determine the canopy coverage and frequency of each plant species. From the data collected in this study, species composition could be determined for each transect of interest. Sampling was conducted on upland and bottomland positions in watersheds 001C, 020B, N20B, N01B, 001D, N04D, and 004B. Only Florence soil upland positions were sampled in watersheds 004D, 004F, 002D, and 002C.

Four transects were established in each upland and lowland position. This research design resulted in the majority of the sampled watersheds containing 8 transects — four on upland and four on bottomland sites. Five 10 square meter circular plots are located on each 50 meter transect at five meter intervals. Sampling is then conducted in each of the five plots along every transect throughout the area.

To assess plant species composition all plant species in the circular plots are recorded. On each plot, canopy coverage of all species are recorded according to the following canopy coverage classes modified from Daubenmire: 1. 0-1%; 2. 1-5%; 3. 5-25%; 4. 25-50%; 5. 50-75%; 6. 75-95%; 7. 95-100% (LTER data set PVC02). The primary investigators identified canopy coverage as the area within the lines

connecting the extremities of the plant canopy represented by a particular species. Species were sampled in mid to late May, mid July, and September. Sites have been sampled on a yearly basis dating back to the early 1980s. This researcher was specifically interested in the species composition data from 1987 as this corresponds to the available imagery.

Despite the fact that all images were supposed to have been completely preprocessed and free from defects, there were several individual scenes which were not acceptable for various reasons. Obvious cloud contamination was evident on 4 of the images. This precluded these images from further analysis. Three of the images were not correctly rectified. This was very apparent when comparison was made with the other images. These problems were an initial setback because this left the entire month of May unrepresented by an image. A simple image to image rectification was performed on a single uncorrected scene from the month of May. The resulting RMS error was less than a pixel.

Seventeen SPOT scenes were eventually deemed acceptable and included in the analysis. These images were imported into Erdas Imagine and overlaid upon each other. This is the process of stacking one image on top of the other. This eventually resulted in a single image containing seventeen

bands – each band representing a different date of the 1987 growing season.

Methods:

It is important to generate temporal trajectories of NDVI for each of the 72 transects if relative amounts of C_3 and C_4 are to be determined based on metrics derived from the curve. Approximate locations of the transects were hand drawn on a field map by the personnel conducting the species composition research. The corresponding locations of these transects on the composited image could be roughly estimated from the information contained in the raster and vector GIS files.

Although an exact agreement between the actual location of the transect and the identified location on the image is somewhat unlikely, the species composition data is believed to be representative of the immediate surrounding area as well. Therefore, it is believed that a subset of the area around each transect will be sufficient for this project. Each of the 72 transects were subset using the Erdas Imagine software. Areas of interest were used to produce temporal trajectories of NDVI for each transect.

The percentage of C_3 and C_4 plants contained within the heterogeneous canopies of the 72 transects located in the

various watersheds is determined using a procedure called discriminate function mixture modeling. Bahdwar (1984) and Goodin and Henebry (1997) have found this to be somewhat effective for distinguishing cover types intermixed in a heterogeneous canopy. These discriminant functions are similar to multiple regressions equations;

$$d = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots \alpha_n X_n$$

Equations similar to these were developed that are based on metrics derived from the cumulative NDVI curve plotted against growing degree day (GDD) for areas of known species composition. Growing Degree Day accumulations involve the amount of accumulated heat in plants and is related to their stage of growth. This is often a better indicator of development than simply time of year due to seasonal variations in temperature (Michigan State Extension Service, 1997). Metrics were selected that will most effectively distinguish between areas consisting of various amounts of C_3 and C_4 plants. For example, an area consisting of 20% C_3 plants would theoretically have a significantly different temporal profile than an area with 80% C_3 plants. Therefore, the Growing Degree Days at which various percentages of NDVI occur was also expected to vary according to percentage of C_3 and C_4. It was found that the differences associated with temporal profiles of transects

containing various amounts of C_3 and C_4 will be distinguishable using the SPOT satellite data.

Initially, the metrics used to summarize individual NDVI curves included GDD at which 20%, 50%, and 90% of normalized integrated NDVI had accumulated. It is possible that, due to the 30 meter spatial resolution of the images, pixel contamination by biotic material not fitting the classes of C_3 and C_4 could occur. This is increasingly likely later in the growing season when C_3 and C_4 plants are active in the study area. This might result in one or more of the metrics being ineffective at distinguishing the percentage class of C_3 and C_4. It appears that this may have occurred in a number of transects used in this study at times late in the growing season.

An additional metric was included in the analysis to try and account for this problem. This metric attempted to summarize the NDVI curve at a point early in the growing season when only a single lifeform type was present. The GDD when 35% of normalized integrated NDVI accumulated was determined. It appears that this improved classification moderately.

Comparisons were then made between classifications produced using various metrics(See Tables 1-5 for results). Method I, II, and III all use metrics derived from the

temporal trajectories of NDVI developed from the composited 17 layer image. In order to explore the effect of lower frequency data on the accuracy of classification, methods IV and V used metrics developed from only 10 images. This resulted in a 10 layer composited image. These temporal profiles were derived using approximately one image per month throughout the length of the growing season.

Method I : GDD at 20%,50%,90%

Method II : GDD at 20%,35%,50%,90%

Method III: GDD at 20%,35%,50%

Method IV : GDD at 20%,50%,90%(1 image/month)

Method V : GDD at 20%,35%,50%(1 image/month)

Integrated NDVI has been found to be a rather useful measurement when analyzing vegetation characteristics using remote sensing. Malingreau notes that "as the NDVI measurement is representative of a rate of photosynthetic plant activity, its integration over time should tell us more about the productive history of the plant" than other methods (Malingreau, 1989). Expression of the integrated NDVI curve can be somewhat difficult and can been estimated by the equation;

$$\int NDVI \approx NDVI \times days = \Sigma\,(NDVI \times n)$$

where 'n' is the number of days in each period covered by an

NDVI measurement.

Growing degree day was calculated using information from the Konza Prairie data server maintained by the Division of Biology. A multitude of daily climatic information is archived dating back to the early 1980s. Growing degree day was calculated using a base temperature of 4 degrees Celsius. The selection of a base temperature is dependant upon the area of investigation. This is the temperature at which growth takes place for vegetation of a particular area. If the air temperature falls below the base temperature then no degree days are accumulated. Goodin and Henebry (1997) have demonstrated the effectiveness of using a base of 4 degrees in nearly identical prairie conditions. Average daily temperature was one of the variables included in the archived data and was also used to calculate GDD. The formula used to calculate GDD is as follows;

$$GDD = \Sigma[(\text{Avg. Daily Temp.}) - 4°C]$$

Temporal trajectories were developed for each of the 72 transects that plotted Cumulative Integrated NDVI against GDD. Percentage of C_3 an C_4 lifeforms contained within each transect was determined by classifying individual species into their appropriate lifeform category. Dominant species found within the transects include C_3 lifeforms *Poa*

pratensis and *Agropyron smithii* and C_4 lifeforms *Andropogon gerardii* and *Sorghastrum nutans*.

The use of discriminant mixture modeling for classification requires that a subset of the data be used to derive classification coefficients. A random selection of 36 of the original 72 transects was used for this purpose. The GDD at which 20%, 35%, 50%, and 90% of cumulative NDVI occurred was determined for each of these transects using a polynomial fitted piecewise to pass through the data. This data was then subjected to a number of multivariate tests including Wilks' Lambda, Pillai Trace, and Hotelling-Lawley Trace. These tests were used to ensure that significant differences existed between each class. If this was not the case, there would have been little point in continuing to develop discriminant scores. Classification coefficients and discriminant scores were calculated for each category once it was determined that the multivariate tests were significant. Discriminant scores for four classes were developed. These include (1)<5% C3, (2)5-25% C3, (3)25-50% C3, (4)51-100% C3.

Results

An examination of the results of the classifications using the various metrics reveals that all methods produced

overall accuracies above 50% (Tables 1-5). All methods were accurate to near the 60% level. Methods II, III, and V produced results with the highest overall levels of accuracy approaching 62%. Methods I and IV were very close with overall accuracies of 59% respectively.

An important point to note about classification accuracies is the fact that even a completely random assignment of cases to a particular class will produce a certain percentage of correct values in the error matrix. In certain instances, random assignments can result in a relatively good apparent classification result. Although the overall accuracy is a useful aid in understanding the general effectiveness of a particular classification technique, it does not indicate how much better this is than a completely random assignment of classes.

The k statistic is a measure of the difference between the level of accuracy of a particular classification and a completely random classification. Lillesand and Kiefer define the k statistic as follows:

$$k = \frac{\text{observed accuracy - chance agreement}}{\text{1 - chance agreement}}$$

Lillesand and Kiefer (1994) note that this statistic "serves as an indicator of the extent to which the percentage correct values of an error matrix are due to true versus chance agreement."

When the k statistic is included in the analysis, more variation can be seen between the techniques. The k statistic ranges from a low of 32% for method II to a high of 44% for method V. In other words, method V produced a classification that is 44% better than random chance. An apparently common trend in the results is that when the metric associated with the GDD at which 35% of cumulative NDVI is reached is included in analysis the accuracy of the classification increases.

Category	0-5%	5-25%	25-50%	>50%
0-5%	5	1	1	
5-25%	7	14	5	
25-50%			2	1
>50%				1
User accuracy:	41.6%	93.3%	25%	50%

Total %correct: 59.5%
K-hat: 36.1%

Table 1: Classification Matrix for Method I

There are several possible reasons why the inclusion of this particular metric improves the results. One explanation relates to the fact that there is a rather small, but still important, percentage of biotic material within the observed canopies that cannot be classified as C_3 or C_4. This could result in the distinct temporal trajectories of C_3 and C_4 becoming mixed, or weakened, by

other vegetative matter. This would likely occur towards the mid to late part of the growing season when all different lifeforms are present in the canopy. Therefore, the indicated GDD at which 90% of NDVI accumulates could be significantly affected by these other kinds of material. However, during the early part of the growing season these other lifeforms would likely not be as advanced and could not cause as much confusion between classes as is apparently the case later in the season.

When the classification of individual categories, or the user's accuracy, is considered, some interesting trends seem to emerge. The accuracy of classes two and four were consistently the highest of the group. Classes 2 and 4 correspond to transects containing between 25% and 50% C_3 and sites containing over 50& C_3. It should be noted however, that there were only a very small number of cases to be categorized in class 4. This was do to the small number of transects of this class in the original data.

Category	0-5%	5-25%	25-50%	>50%
0-5%	4			
5-25%	8	15	6	
25-50%				
>50%			2	2
User accuracy:	33%	100%	0%	100%

Table 2: Classification Matrix for Method II

Category	0-5%	5-25%	25-50%	>50%
0-5%	9	4	2	
5-25%	1	11	4	
25-50%	2		1	
>50%			1	2
User accuracy:	75.0%	73.3%	12.5%	100%

Total %correct: 62.2%
K-hat: 43.6%

Table 3: Classification Matrix for Method III

Although not as accurate as classes 2 and 4, class 1 generally achieved a respectable user's accuracy. These ranged from a low of 33% using method II to a high of 75% using method III. Cases that belonged to this class were most often incorrectly classified as belonging to class 2. This is not necessarily a surprise as many of the cases classified were borderline between class 1 and 2.

Class 3 was routinely incorrectly classified using all methods but number IV. Method IV yielded a user's accuracy of 87.5% for this particular category. No other method yielded results higher than 25% for this class. The reasons for this are not clear at this time. However, the temporal trajectories used in the classification were effectively smoothed by using fewer images in the analysis as compared to other methods. Method V, though, which utilized the same number of images did not achieve a user's accuracy as high

for class 3.

Category	0-5%	5-25%	25-50%	>50%
0-5%	8	5	1	
5-25%	4	6		
25-50%		4	7	1
>50%				1
User accuracy:	66.7%	40%	87.5%	50%

Total %correct: 59.5%
K-hat: 41.7%

Table 4: Classification Matrix for Method IV

Category	0-5%	5-25%	25-50%	>50%
0-5%	8	2		
5-25%	4	11	4	
25-50%		1	2	
>50%		1	2	2
User accuracy:	66.7%	73.3%	25%	100%

Total %correct: 62.2%
K-hat: 44.3%

Table 5: Classification Matrix for Method V

Although the results are somewhat variable and trends are difficult to establish, all methods classified the cases into their appropriate categories better than simple random chance. The majority of cases that were incorrectly classified were assigned to a contiguous class. This is in agreement with what Goodin and Henebry found in their

research. This seems to be a promising indication that this technique is able to detect differences in the temporal trajectories of the NDVI curve and relate them to percentage C_3 lifeform present.

It appears from this limited research that no significant degradation of accuracy results when the number of images used for analysis is decreased. In fact, the overall accuracy, *k* statistic, and user's accuracy are all quite high with metrics derived from lower frequency data. The inclusion of the metric associated with GDD at 35% appears to improve classification slightly. This may be due to the reasons already discussed. However, more research would obviously need to be done on this matter.

Conclusions

The purpose of this report was to determine the relative amounts of C_3 and C_4 lifeforms present in a heterogeneous canopy. Goodin and Henebry have shown that this is possible using close range remote sensing measurements of fine spatial resolution. However, the ability to distinguish lifeforms contained within a 30 meter pixel was unclear. This report has demonstrated that this is somewhat possible using data with a coarser spatial resolution. Metrics based on the cumulative NDVI curve and

GDD were able to classify transects with unknown membership into the appropriate categories the majority of the time. Although the accuracy of the classifications were only in the 60% range, this is respectable considering the spectral similarity of the lifeforms at any single point in time. This is also significant given the 30 meter spatial resolution of the images used in analysis, and the associated amount of biotic material not fitting the classes of C_3 or C_4 which was undoubtedly contained in some of the pixels.

Transects that contained little or no material other than C_3 and C_4 were generally more often classified correctly than those sites with additional biotic material not fitting these classes. It is not clear how much of these additional materials need to be present in a particular pixel in order for classification accuracy to suffer. However, there does appear to be a link between classification accuracy and the amount of biotic matter that is neither C_3 or C_4.

Another factor which could have caused some analytical problems is the quality of the species composition data that were used to generate classification coefficients for the discriminant analysis, and how closely this represented the actual situation on the ground. Although these data were

collected with supervision from the Division of Biology, there are no known quality control measures that were undertaken to ensure the data is accurate. Furthermore, the fact that the transects subset from the composited image only approximated the actual collection transects on the ground could cause some additional inconsistencies in the results.

A possible way to improve on the uncertainties associated with the location of the transects is to utilize the innovative technology of the Global Positioning System (GPS). A GPS could be taken to the field and exact boundaries of the transects could be determined. Unfortunately, this information was not available for this project due to certain limitations. However, the geocoding of all transect locations using GPS is currently underway.

This report has demonstrated the effectiveness of using satellite images and metrics based on temporal trajectories of NDVI to determine relative amounts of C_3 and C_4 photosynthetic pathways. The ability to distinguish these lifeforms using satellite imagery is important if this is to become an effective technique for monitoring the condition of these environments on a widespread scale. The results presented here seem a to be a promising indicator that this is possible.

References

Abrams, M.D., and Hulbert, L.C., (1986), Effect of topographic position and fire on species composition in tallgrass prairie in northeast Kansas, *The American Midland Naturalist*, 117(2): 442-445.

Althausen, J.D., Narumalani, S., Jenson, J., Burkhalter, Shan., and Mackey, H.E. Jr., (1997), Aquatic macrophyte modeling using GIS and logistic multiple regression, *Photogrammetric Engineering and Remote Sensing*, 63(1):41-48.

Bidwell, T.G., Engle, D.M., and Claypool, L.P., (1980), Effects of spring headfires and backfires on tallgrass prairie, *Journal of Range Management*, 43(3): 209-216.

Collins, S.L., (1992), Fire frequency and community heterogeneity in tallgrass prairie vegetation, *Ecology*, 73(6):2001-2006.

DeFries, R., Hansen, M. and Townshend, J., (1995), Global discrimination of land cover types from metrics derived from AVHRR pathfinder data, *Remote Sensing of Environment*, 54(3):209-222.

Fischer, A., (1994), A model for the seasonal variations of vegetation indices in course resolution data and its inversion to extract crop parameters, *Remote Sensing of Environment*, 48: 220-230.

Fleischmann, C.G., and Walsh, S.J., (1991), Multi-temporal AVHRR digital data: an approach for landcover mapping of heterogeneous landscapes, *Geocarto International*, 4:5-20.

Gibson, D.J., and Hulbert, L.C., (1987), Effects of fire, topography, and year to year climatic variation on species composition in tallgrass prairie, *Vegetatio*, 72:175-185.

Goodin, D.G., and Henebry, G.M. A technique for monitoring ecological disturbance in tallgrass prairie using seasonal NDVI trajectories and a discriminant function mixture model (publication forthcoming in *Remote Sensing of Environment,* 1997).

Henebry, G.M., (1993), Detecting change in grasslands using measures of spatial dependence with landsat TM data, *Remote Sensing of Environment*, 46:223-234.

Kiefer, Ralph, and Lillesand, Thomas. *Remote Sensing and Image Interpretation*. New York: John Wiley & Sons, 1994.

Klecka, William, R., *Discriminant Analysis*. London: Sage Publications, 1980.

Kremer, R.G., and Running, S.W., (1993), Community type differentiation using NOAA/AVHRR data within a sagebrush-steppe ecosystem, *Remote Sensing of Environment*, 46:311-318.

Lascassies, P.P., Flouzat, G., Gay, M., and Vignolles, C., (1994), Validation of the use of multiple linear regression as a tool for unmixing coarse spatial resolution images, *Remote Sensing of Environment*, 49:155-166.

Lewis, J.K., and Waller, S.S., (1979), Occurrence of C3 and C4 photosynthetic pathways in North American grasses, *Journal of Range Management*, 32(1):12-28.

Malingreau, J.P., (1989), The vegetation index and the study of vegetation dynamics, *Applications of Remote Sensing to Agrometeorology*, 285-303.

Michigan State Extension Service, Growing degree day accumulations, <www.msue.msu.edu/msue/imp/mod03/03900056.html>, Last modified 1997.

Reed, B.C., Brown, J.F., VanderZee, D., Loveland, T.R., Merchant, J.W., Ohlen, D.O., (1994), Measuring phenological variability from satellite imagery, *Journal of Vegetation Science*, 5(5): 703-714.

Towne, G., and Owensby, C., (1984), Long term effects of annual burning at different dates in ungrazed Kansas tallgrass prairie, *Journal of Range Management, 37(5): 392-399.*

Zhuang, H.C., Shapiro, M., and Bagley, C.F., (1993), Relaxation vegetation index in non-linear modelling of groung plant cover by satellite remote sensing data, *International Journal of Remote Sensing,* 14(18):3447-3470.